


# Flow cytometric monitoring of bacterioplankton phenotypic diversity predicts high population-specific feeding rates by invasive dreissenid mussels

Ruben Props<sup>1,2,3</sup> ,<sup>1,2,3</sup> Marian L. Schmidt,<sup>3</sup>  
Jasmine Heyse,<sup>1</sup> Henry A. Vanderploeg,<sup>4</sup>  
Nico Boon<sup>1</sup> and Vincent J. Denef<sup>3\*</sup>

<sup>1</sup>Center for Microbial Ecology and Technology (CMET),  
Ghent University, Coupure Links 653, 9000 Gent,  
Belgium.

<sup>2</sup>Belgian Nuclear Research Centre (SCK•CEN),  
Boeretang 200, 2400 Mol, Belgium.

<sup>3</sup>Department of Ecology and Evolutionary Biology,  
University of Michigan, Ann Arbor, MI, USA.

<sup>4</sup>NOAA Great Lakes Environmental Research  
Laboratory, Ann Arbor, MI, USA.

## Summary

**Species invasion is an important disturbance to ecosystems worldwide, yet knowledge about the impacts of invasive species on bacterial communities remains sparse. Using a novel approach, we simultaneously detected phenotypic and derived taxonomic change in a natural bacterioplankton community when subjected to feeding pressure by quagga mussels, a widespread aquatic invasive species. We detected a significant decrease in diversity within 1 h of feeding and a total diversity loss of  $11.6 \pm 4.1\%$  after 3 h. This loss of microbial diversity was caused by the selective removal of high nucleic acid populations ( $29 \pm 5\%$  after 3 h). We were able to track the community diversity at high temporal resolution by calculating phenotypic diversity estimates from flow cytometry (FCM) data of minute amounts of sample. Through parallel FCM and 16S rRNA gene amplicon sequencing analysis of environments spanning a broad diversity range, we showed that the two approaches resulted in highly correlated diversity measures and captured the same seasonal and lake-specific patterns in community composition. Based on our results, we predict that selective feeding by**

**invasive dreissenid mussels directly impacts the microbial component of the carbon cycle, as it may drive bacterioplankton communities toward less diverse and potentially less productive states.**

## Introduction

Anthropogenic disturbances can lead to rapid changes in microbial community diversity (species richness, evenness and composition). Many studies aim to better understand feedbacks between global change and microbial communities, as changes in microbial diversity can either mitigate the predicted direct effects of disturbances on ecosystem fluxes (Singh *et al.*, 2010; Zhou *et al.*, 2012), or lead to major shifts in bacterially mediated fluxes (Schimel and Gullede, 1998; Finlay *et al.*, 2007; Levine *et al.*, 2011). The responses of microbial communities to disturbances are often monitored by means of high-throughput molecular techniques, such as 16S rRNA gene amplicon sequencing (Shade *et al.*, 2012). Community shifts in response to altering environmental parameters can occur within hours (Props *et al.*, 2016b) to days (Datta *et al.*, 2016) and demand substantial sampling effort at a preferably fixed frequency to allow accurate statistical inference (Faust *et al.*, 2015). Current technology allows sequencing data to be generated from low-volume samples (e.g., 100 ml) of low-density environments ( $\leq 10^6$  cells ml<sup>-1</sup>), which comprise many aquatic ecosystems, but larger sample volumes ( $> 1$  l) are required in order to yield a robust census of the microbial community (Padilla *et al.*, 2015). This prohibits the use of this approach in many longitudinal microcosm studies, for which repeated invasive sampling itself would act as a disturbance.

Recently, a new approach has been developed that can generate phenotypic diversity metrics based on physiological information derived from flow cytometry (FCM) data (Props *et al.*, 2016a). These diversity metrics have been shown to be highly correlated to taxonomic diversity, as derived from amplicon sequencing. Yet, their derivation avoids invasive, high volume sampling practices ( $\leq 1$  ml of sample required) and simultaneously offers information on the physiological state of the community, as well as on the

Received 10 February, 2017; accepted 30 September, 2017. \*For correspondence. E-mail vdenef@umich.edu; Tel. (+1) 734 764 6481; Fax (+1) 734 763 0544.

absolute density of its constituent populations. Briefly, this approach performs kernel density estimations on multiple bivariate single-cell parameter combinations (e.g., fluorescence and scatter intensity) and concatenates these into a feature vector that is called the *phenotypic fingerprint*. The phenotypic fingerprint represents the community structure in terms of physiological aspects, such as nucleic acid content and morphology. From this fingerprint, the community diversity can be calculated by means of the Hill diversity numbers (Hill, 1973), which examine both richness and evenness components of the phenotypic community structure. In parallel, this approach facilitates beta-diversity assessments through the ordination of samples by means of a dissimilarity matrix calculated between phenotypic fingerprints. The ability to simultaneously track impacts on phenotypic and taxonomic diversity offers opportunities to address gaps in our understanding of microbial disturbance ecology. Currently, this method has only been tested in one, low-complexity system and validation across a broader range of diversities is needed to fully assess its potential.

Species invasion, which is one of the main components of global change (Chapin *et al.*, 2000), is a particularly useful system to help address knowledge gaps in microbial disturbance ecology as we can readily mimic the real-world conditions (i.e., sudden introduction) in laboratory or field experiments. The current distribution of invasive dreissenid mussels (IDMs) across North America (> 30 states) is a prime example of a successful invasion event (Higgins and Vander Zanden, 2010). Initially introduced through ballast water, IDMs display high filtration rates (Vanderploeg *et al.*, 2002) and are able to rapidly populate benthic and littoral substrates in densities of up to 19,000 individuals per m<sup>2</sup> (Nalepa *et al.*, 2010). With respect to their feeding behaviour, IDMs show highly selective feeding behaviour toward seston and different algal and microzooplankton taxa over a broad range of size (~ 1–200 µm) (Tang *et al.*, 2014). While IDMs are known to strongly impact phytoplankton and zooplankton abundance and composition (Higgins and Vander Zanden, 2010), the few studies focused on their impacts on bacterioplankton report contradicting results. Several of these studies reported selective feeding on bacterial species (Silverman *et al.*, 1995; Pires *et al.*, 2004; Denef *et al.*, 2017), while a long-term environmental survey of the Hudson River prior- and post-invasion did not observe negative effects on bacterial community density and productivity (Findlay *et al.*, 1998).

In this study, we investigated the effect of IDM grazing (with *Dreissena bugensis* as model) on the natural bacterioplankton community of Lake Michigan through (near) non-invasive tracking of the phenotypic biodiversity, as well as the density of physiological subpopulations. We first validated whether the existing correlation between taxonomic and phenotypic diversity metrics holds for the high diversity

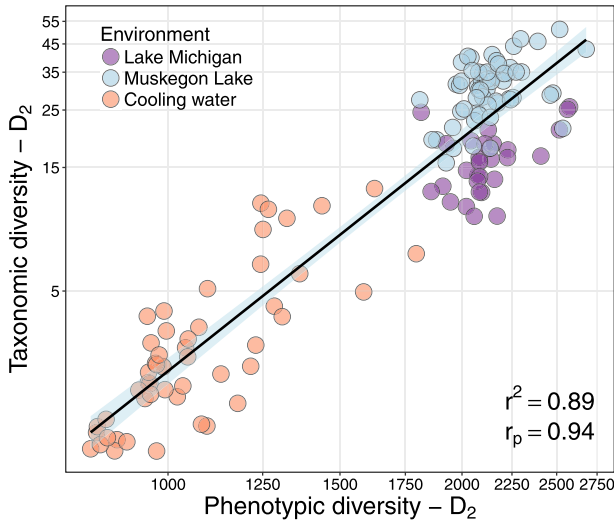
environments of Lake Michigan (low primary and secondary productivity) and one of its freshwater estuaries (high primary and secondary productivity), Muskegon Lake. We then used phenotypic alpha and beta diversity analyses to assess the impact and extent of IDM grazing on the bacterioplankton community of Lake Michigan. The observed biodiversity dynamics were further related to the dynamics of well-established physiological populations in freshwater bacterioplankton, for which the grazing rate was determined.

## Results

We aimed to use FCM-derived phenotypic diversity metrics as a proxy for taxonomic diversity shifts occurring during quagga mussel feeding on lake bacterioplankton. The experiment consisted of 12 l microcosms of 153 µm screened water and quagga mussels, both retrieved from Lake Michigan. The Lake Michigan bacterioplankton has previously been shown to contain both grazing-resistant and grazing-sensitive taxa allowing the study of direct grazing impacts (Tang *et al.*, 2014; Denef *et al.*, 2017). Prior to the onset of the experiment, we assessed whether the previously established correlation between phenotypic and taxonomic diversity metrics for low diversity environments could be extended to higher diversity aquatic environments, such as Lake Michigan.

### *Validation of phenotypic diversity as a proxy for taxonomic diversity*

Microbial communities can be classified as relatively low or high diverse communities based on their Hill diversity metrics as these are expressed in terms of *effective number of taxa*, which depict the number of equally abundant taxa required to obtain the same diversity value as the community in question (Hill, 1973). The diversity metrics derived from flow cytometric analysis are calculated in the same way as their taxonomic counterparts, but they can only be interpreted in arbitrary units. In order to determine whether there was a general relationship between the phenotypic and taxonomic diversity, we included an independent dataset from a low diversity cooling water microbial community (1–10 effective number of taxa) (Fig. 1). This cooling water dataset contains two 40-day surveys of bacterioplankton communities present in the secondary cooling water system of a nuclear test reactor that was subjected to multiple operational phases. The positive correlation between taxonomic and phenotypic diversity metrics of orders 1 ( $D_1$ , exponential of Shannon entropy) and 2 ( $D_2$ , inverse Simpson index) could be extended to the Lake Michigan and Muskegon Lake communities (cross-validated  $r^2 = 0.89$ , Pearson's correlation  $r_p = 0.94$  for both  $D_1$  and  $D_2$ ). We opted for a single regression model as opposed to individual regressions for each environment in order to avoid

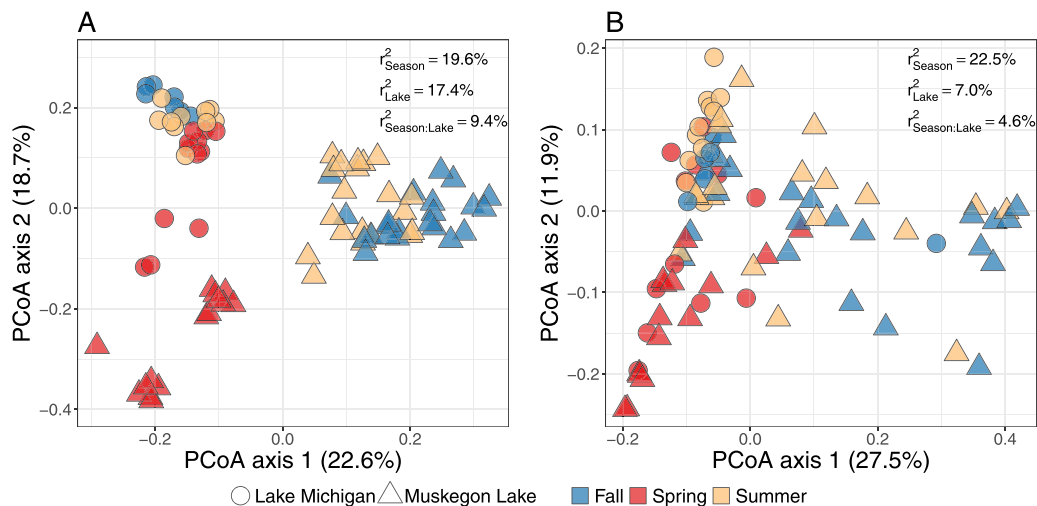


**Fig. 1.** Validation for the use of the phenotypic diversity (derived from FCM) across environments with varying degrees of taxonomic diversity (derived from 16S rRNA gene amplicon sequencing,  $n = 138$ ). The cooling water samples represent bacterioplankton communities sampled throughout two 40-day temporal surveys of a cooling water system of a nuclear test reactor (Props *et al.*, 2016a). Lake Michigan and Muskegon lakes samples represent bacterioplankton communities sampled over a productivity gradient, at various depths (110, 45 and 15 m) and throughout three seasons (Fall, Spring and Summer). Fall, Spring and Summer denote samples taken in September, April and July respectively. The shaded area represents the 95% confidence interval around the ordinary least squares regression model. Both diversities are depicted on a  $\log_2$  scale. In addition to the average variance explained ( $r^2$ ) after tenfold cross validation with 100 repeats, Pearson's correlation coefficient ( $r_p$ ) is also provided. Bootstrap error intervals fell within the label size and were not displayed.

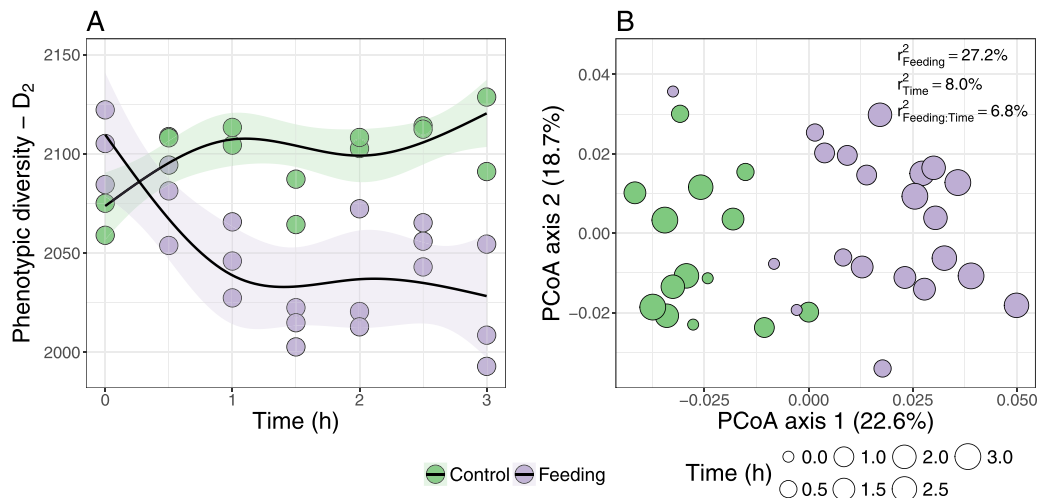
overfitting as well as to construct a generalizable model that provides robust inference for all three environments. Individual regression models did not significantly differ in slope, and the cross-validated  $r^2$  of the single regression model was high ( $r^2 = 0.89$ ), thereby permitting the use of a single regression model for monitoring diversity dynamics (Supporting Information Fig. S1 and Supporting Information Table S1).

The dynamic range of the regression, defined as the ratio between the largest and smallest taxonomic diversity used in its calculation, was 88.7 for  $D_1$  and 42.5 for  $D_2$ . Goodness-of-fit analysis of the linear regression model indicated a normal distributed residual distribution with homogenous variance over the entire regression range (Supporting Information Fig. S2). The observed richness ( $D_0$ ) did not show a distinct linear correlation ( $r^2 = 0.32$ ,  $r_p = 0.54$ , Supporting Information Fig. S3). Due to the high level of correlation between the phenotypic diversity ( $D_1$  and  $D_2$ ) and the taxonomic diversity, it was permissible to use the phenotypic diversity as a stand-alone metric for evaluating bacterioplankton diversity. Additionally, only  $D_2$  was used in further analyses due to the high degree of correlation between  $D_1$  and  $D_2$  ( $r_p = 0.99$ ).

In contrast to alpha diversity, beta diversity cannot be captured by a single metric. Therefore, we compared the taxonomic and phenotypic beta diversity by their performance to detect seasonal- and lake-specific community structures in the Lake Michigan and Muskegon Lake data set (Fig. 2). Procrustes analysis demonstrated that both approaches were significantly correlated in terms of the



**Fig. 2.** Application of the taxonomic (A) and phenotypic (B) beta-diversity (PCoA of Bray–Curtis dissimilarity matrix) to investigate season- and lake-specific effects on the community structure of Lake Michigan and Muskegon Lake ( $n = 87$ ). Fall, Spring and Summer denote samples taken in September, April and July respectively. The variance explained by the lake and season variables, as well as the interaction effect between the lake and season variable is provided (PERMANOVA). All effects were significant at the  $p = 0.001$  level with the exception of the interaction effect for the phenotypic beta diversity ( $p = 0.018$ ). Procrustes analysis confirmed the high degree of correlation between both beta-diversity analyses ( $p = 0.001$ , 999 permutations). Permutations for PERMANOVA and Procrustes analyses were constrained within each survey year.



**Fig. 3.** Feeding effect on the phenotypic alpha diversity (A) and phenotypic beta diversity (B) of the bacterioplankton community. Bootstrap error intervals on the phenotypic diversity were calculated on three technical replicates for each microcosm but fell within the label size and are therefore not displayed. Shaded areas indicate 95% confidence intervals on the robust smoothing spline regressions. Label radius of the data points in the beta-diversity analysis is proportional to the time into the experiment. The variance explained by the overall temporal and feeding effect, as well as the interaction effect between the feeding and experiment time is provided (PERMANOVA). All effects were significant at the  $p = 0.01$  level.

patterns that they captured in the data ( $p = 0.001$ ). Additionally, both approaches identified season and lake type (Lake Michigan or Muskegon Lake) as significant predictors of the community structure ( $p = 0.001$ ). The season explained the most variance in the beta diversity (i.e., 19.6% of the taxonomic beta diversity and 22.5% of the phenotypic beta diversity). Lake type still captured 17.4% of the variance in the taxonomic beta diversity, but only 7.0% in the phenotypic beta diversity. Finally, the seasonal effect was dependent on the lake type, representing an extra 9.4% of the variance in the taxonomic beta diversity ( $p = 0.001$ ) and 4.6% in the phenotypic beta diversity ( $p = 0.018$ ).

#### Diversity dynamics during IDM feeding

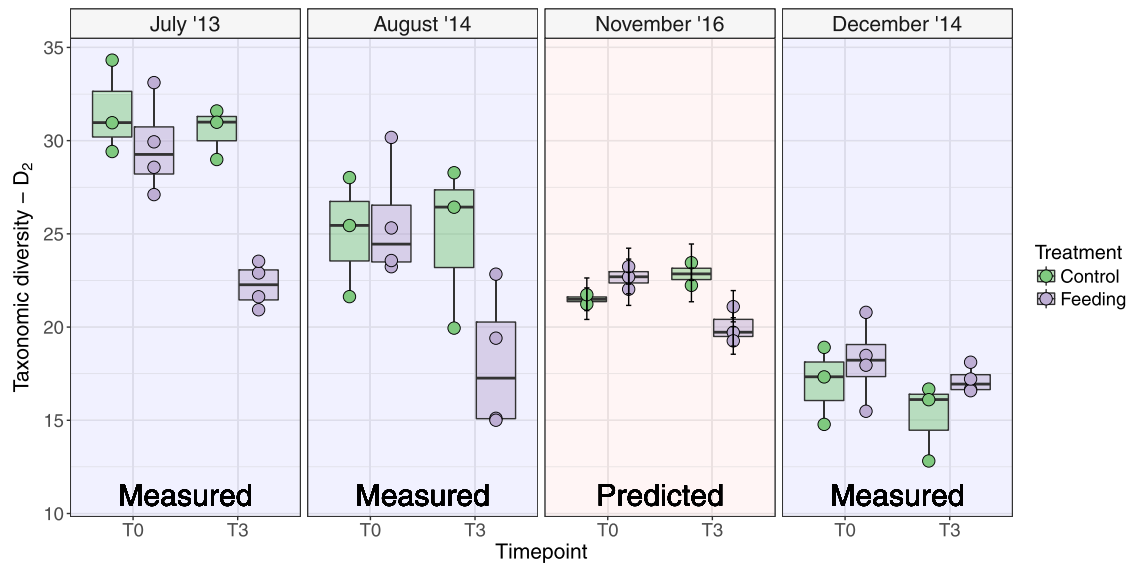
The temporal trajectory of the bacterioplankton community of the microcosms was monitored for 3 h at a resolution of 0.5 h when subjected to the direct feeding pressure by 15 IDMs per microcosm (Fig. 3). This time period was sufficiently long to allow the assessment of direct feeding effects (removal of 30–60% of seston), but short enough to avoid indirect effects, e.g., due to trophic cascades or substantial accumulation of feces and pseudofeces (Vanderploeg *et al.*, 2010). Importantly, all mussels were subjected to an extensive pretreatment consisting of specific handling, rinsing and acclimatization steps in order to avoid contamination of the bacterioplankton community by external periphyton, debris and ingested particles at the onset of the experiment (see *Experimental procedures* section). The size distribution of the mussels

was not significantly different between the microcosms ( $22.7 \pm 2.3$  mm, Kruskal–Wallis test,  $p = 0.08$ ). The total dry weight of the mussels per microcosm was  $0.24 \pm 0.018$  g DW.

Over the span of the experiment, the control microcosm's phenotypic diversity exhibited a minor overall positive temporal drift ( $p = 0.038$ ). In contrast, the bacterioplankton phenotypic diversity underwent a clear and significant decrease ( $p < 0.0001$ ) during filter feeding of the IDMs, signifying the enrichment of the community by certain taxa (Fig. 3A). The treatment effect became significantly distinguishable from the control at the 1 h mark (at  $p < 0.05$ ). Using the regression model, an average loss in taxonomic diversity ( $D_2$ ) could be predicted of  $2.6 \pm 1.0$  effective number of taxa, corresponding to a decrease of  $11.6 \pm 4.1\%$  over the course of the experiment. Conceptually, this means that in a hypothetical community of 23 equally abundant taxa (diversity prior to mussel feeding), an average of 2.6 taxa would be lost due to IDM feeding. To put these measurements in perspective, we analysed data from a recently published mussel-feeding study that used the same experimental design and had 16S rRNA gene amplicon data at time points 0 and 3 h available (Fig. 4) (Denef *et al.*, 2017). We calculated a mean loss of taxonomic diversity ( $D_2$ ) of  $5.32 \pm 4.65$  for their three independent experiments, which is comparable to the taxonomic diversity loss predicted for our experiment ( $2.6 \pm 1.0$ ).

While the monitoring of the phenotypic alpha diversity allowed us to track the treatment effect through time, a beta diversity analysis was also conducted to evaluate the





**Fig. 4.** Measured or predicted taxonomic alpha diversity of Lake Michigan bacterioplankton communities after a three hour exposure to invasive quagga mussels. Measured taxonomic diversity data is publicly available from Deneff and colleagues (2017). The measured data comes from feeding experiments that were carried out with Lake Michigan bacterioplankton communities retrieved over a two year period and under the identical experimental design as described in this manuscript. Predicted alpha diversity values were calculated based on the phenotypic diversity data generated in this study.

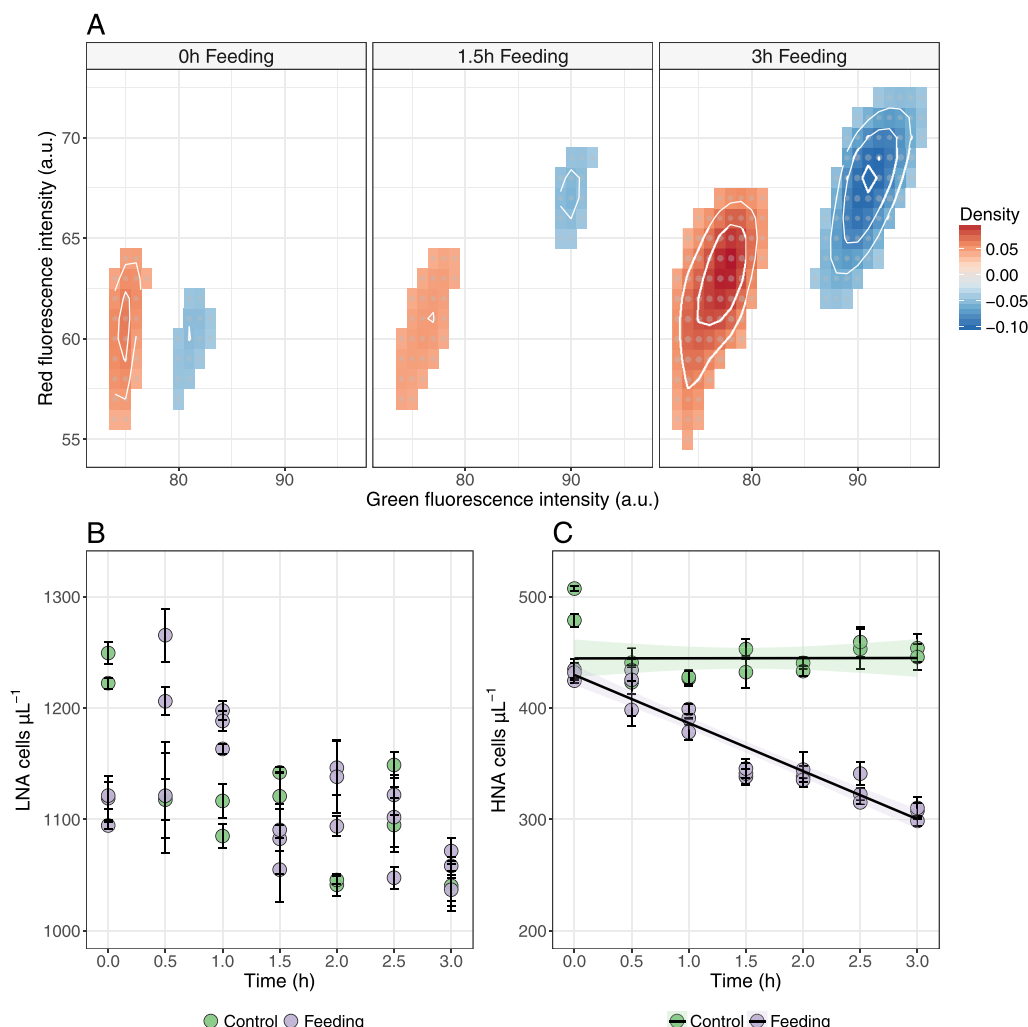
treatment and temporal effects on the complete phenotypic structure of the bacterioplankton community (Fig. 3B). Over the course of the experiment the bacterioplankton communities of the control and treatment microcosms became more dissimilar. In agreement with the alpha diversity analysis, a time-dependent treatment effect ( $r^2_{\text{feeding}} = 0.34$ ,  $p = 0.005$ ) was driving the bacterioplankton structure, with the control bacterioplankton community also experiencing a minor temporal effect ( $r^2_{\text{control}} = 0.08$ ,  $p = 0.006$ ).

#### *Bacterioplankton population dynamics during IDM feeding*

Next, we investigated whether the observed diversity dynamics were caused by selective feeding on specific phenotypic populations of the bacterioplankton community. To do so, contrasts between the phenotypic fingerprints of the treatment and the control at three different time points were calculated (Fig. 5A). This analysis allowed the visualization of regions in a specified bivariate parameter space which are relatively more or less abundant in the treatment versus the control. We opted for the primary fluorescence channels of the SYBR Green nucleic acid stain (i.e., FL1-H and FL3-H) allowing us to identify distinct physiological populations with varying nucleic acid content (Gasol *et al.*, 1999; Hammes and Egli, 2010; Koch *et al.*, 2014). The results demonstrate that during filter feeding the

bacterioplankton community became enriched with a low nucleic acid content population (LNA, low FL1-H/FL3-H intensity) and was depleted from a high nucleic acid (HNA) content population (high FL1-H/FL3-H intensity).

As these are relative changes that do not necessarily reflect a direct feeding effect on the HNA population, the absolute abundances for both the HNA and LNA population were extracted from the total community according to the guidelines by (Prest *et al.*, 2013) (Supporting Information Fig. S4). The LNA cell densities show similar temporal behaviour for the control (coefficient of variation [CV] = 5.8%) and treatment (CV = 5.2%) microcosms (Fig. 5B). This level of variation falls within the technical variation of current FCM technology (CV = 5%) and is thus not indicative of a feeding effect (Hammes *et al.*, 2008). In contrast, the HNA population was directly affected by the filter feeding (Fig. 5C). The HNA population of the control microcosms displayed a similar variation to the LNA population (CV = 5.1%), while the HNA population in the treatment microcosms showed a monotonic decrease throughout the experiment (CV = 12.8%). Analogous to the diversity analyses, a significant treatment effect could be detected within 1 h. Using robust linear regression, the HNA-specific removal rate was estimated at  $43,000 \pm 3000 \text{ cells ml}^{-1} \text{ h}^{-1}$  ( $p < 0.0001$ ), while the control HNA cell density remained constant ( $p = 0.98$ ). After 3 h of being subjected to filter feeding,  $29 \pm 5\%$  of the HNA bacterioplankton population was removed from



**Fig. 5.** Dynamics of high nucleic acid (HNA) and low nucleic acid (LNA) populations.

A. Contrasts between the flow cytometric fingerprints of the control samples and the feeding samples after 0, 1.5 and 3 h. Red contours indicate an increase in the LNA population density during feeding, while blue contours indicate a decrease in the HNA population density during feeding, both of which are relative to the control samples at the specified time point. Only contrasts with densities  $> 10.041$  were visualized.

B. Cell density of the LNA population over the course of the feeding experiment. Error bars indicate standard deviation across technical replicates ( $n = 3$ ).

C. Cell density of the HNA population over the course of the feeding experiment. Error bars indicate standard deviations across technical replicates ( $n = 3$ ). Shaded areas indicate 95% confidence intervals on the robust linear regression models.

the water column. The clearance rate on the HNA population was  $4.56 \pm 0.81 \text{ ml mg}^{-1} \text{ DW h}^{-1}$ .

## Discussion

Our understanding of microbial disturbance ecology has been partially constrained by a lack of temporal resolution, caused by methodological limitations in either sampling, logistics or analysis. In order to combat these bottlenecks, we further developed and validated a (near) non-invasive FCM-based approach dedicated to detect changes in the phenotypic diversity of microbial communities. We validated

and applied these phenotypic diversity metrics to natural, high diversity environments, and investigated the response of bacterioplankton communities to a filter feeding disturbance caused by IDMs, which are highly sensitive to the invasive sampling imposed by alternative monitoring techniques (i.e., they would cease their filter feeding activity). Our experimental results highlight the sensitivity of our method to detect subtle changes in diversity over short timeframes. Based on (i) our presented relationship between phenotypic and taxonomic diversity, (ii) the positive relationship that exists between HNA bacterial populations and bacterial productivity in many ecosystems (Zubkov *et al.*, 2001;

Servais *et al.*, 2003) and (iii) the results of our filter feeding experiment, we hypothesize that IDM feeding directly influences both microbial diversity and ecosystem functionality.

#### *Relation between taxonomic and phenotypic diversity*

The regression between the taxonomic and phenotypic diversity data is in agreement with and expands upon previous research (Props *et al.*, 2016a) and offers further insight into the fundamental relation between these metrics. Concretely, the regression's dynamic range has been extended from 10.3 in the previous study (cooling water) to 42.5 for the diversity of order 2 ( $D_2$ ). The quality of the cross-validated regression is worth highlighting as there were substantial differences in the sample treatment and data generation of the data sets, which could have led to systematic bias. First, the FCM samples of Lake Michigan and Muskegon Lake were fixed with glutaraldehyde and archived at  $-80^{\circ}\text{C}$ , whereas the cooling water samples were analysed directly. The glutaraldehyde fixative used in this research has been shown to increase autofluorescence and may have increased the instrument noise (Gunther *et al.*, 2008). Nevertheless, the bacterioplankton community could be reproducibly isolated from the raw data with one fixed denoising strategy for the entire data set (Supporting Information Fig. S4). Second, the amplicon sequencing of the 16S rRNA gene targeted the V4 region for the Lake Michigan and Muskegon samples, whereas the V3–V4 region was targeted for the cooling water samples. This difference in sequenced region has been shown to potentially alter the observed diversity (Schmalenberger *et al.*, 2001; Yu and Morrison, 2004). We did not observe this bias in the taxonomic diversity, but we did observe that the phenotypic diversity was incapable of resolving the Lake Michigan and Muskegon Lake communities, even though they had distinct taxonomic diversities. The regression analysis showed that the strength of the relationship between the phenotypic and taxonomic diversity was unaffected, but that the intercept of the linear models was different ( $p < 0.001$ ). Therefore, we emphasize that the absolute values of the phenotypic diversity metrics need to be compared within a single ecosystem or experimental setting. Finally, rescaling or subsampling the community composition to an equal library size, which is a common yet debated practice in microbial community analyses (McMurdie and Holmes, 2014), did not negatively affect the regression (Supporting Information Fig. S5). In fact, it improved the regression for all diversity metrics and in particular  $D_0$ , indicating that the phenotypic diversity metrics are primarily sensitive to fluctuations in the density of abundant taxa. As such, we recommend to evaluate changes in richness ( $D_0$ ) solely by means of 16S rRNA gene amplicon sequencing, for which novel statistical approaches are being developed that account for both observed and

unobserved taxa (Willis and Bunge, 2015). It is important to note that in order to compare FCM-derived diversity metrics with each other, the underlying raw data must have been generated by the same flow cytometer platform with identical detector and flow rate settings, which was the case for all data presented in this study.

The linearity on the log-scale implies that the change in phenotypic diversity required to detect a corresponding change in taxonomic diversity systematically increases (Fig. 1). For example, at low diversities, a change from 1000 to 1500 units in phenotypic diversity corresponds to a predicted change of 3.4 units in the taxonomic diversity, while an increase from 1500 to 2000 units only corresponds to a predicted change of 2.0 units in the taxonomic diversity. This is one of the limitations of relying on a fixed number of phenotypic parameters (i.e., fluorescence and scatter intensities); the available parameter space that bacterial cells can occupy is limited, resulting in a loss of sensitivity at higher diversities.

The beta diversity analyses yielded similar statistical inference on the seasonal- and lake-specific effects, with both the taxonomic and phenotypic beta diversity identifying seasonality as the most important predictor of community structure. A higher degree of variance could be explained by the lake type in the taxonomic beta diversity, which suggests that the phenotypic approach was less sensitive to differences in taxon distributions between the lakes or that additional variation based on phenotypic plasticity weakens the relationship between taxonomic and phenotypic beta diversity. This is congruent with the alpha diversity measurements where Lake Michigan and Muskegon Lake samples showed similar phenotypic diversity despite possessing distinct taxonomic diversities. Overall, Procrustes analysis confirmed that the phenotypic beta diversity was able to largely capture the same patterns in the data as the taxonomic beta diversity. As such, phenotypic beta diversity analyses constitute a valid approach for hypothesis testing in high diversity environments, but they are susceptible to a higher degree of variability and thus generate potentially lower effect sizes (e.g., for the lake type in this analysis).

#### *Diversity and population dynamics*

The phenotypic diversity dynamics during the 3 h filter feeding experiment were more subtle than in our previous study on the cooling water dataset ( $< 150$  vs.  $> 500$  units; Props *et al.*, 2016a) but occurred over a much shorter time scale (3 h vs. multiple days). Importantly, the predicted loss in taxonomic diversity based on the phenotypic diversity ( $2.6 \pm 1.0$ ) lies well within the range of expected diversity shifts ( $5.32 \pm 4.65$ ) calculated from previous experiments which had start and endpoint measurements of community composition (Fig. 5). Our predictions also

suggest (i) a general season-dependent bacterioplankton diversity with a higher diversity in summer than in fall and winter and (ii) a season-dependent feeding effect resulting in a higher diversity loss in summer than in fall and winter. Overall, the conformity of our predictions to these previous experiments further validates the phenotypic diversity approach.

The diversity dynamics suggested that *D. bugensis* was selectively feeding upon a fraction of the bacterioplankton community, thereby altering the community composition and lowering the diversity. This was confirmed by identifying populations that were selectively enriched through contrast analysis, which demonstrated the selective feeding on bacteria with HNA content (HNA bacteria) (Fig. 4A). The HNA clearance rate ( $4.56 \pm 0.81 \text{ ml mg}^{-1} \text{ DW h}^{-1}$ ), which can be interpreted as the water volume that is fully depleted of HNA bacteria per hour, was comparable to those previously reported for *Dreissena polymorpha* feeding on laboratory strains ( $3.5\text{--}4.8 \text{ ml mg}^{-1} \text{ DW h}^{-1}$ ) ranging in size between 1 and 4  $\mu\text{m}$  in length (Silverman *et al.*, 1995). The clearance rates on laboratory strains were measured for high cell densities ( $> 10^7 \text{ cells ml}^{-1}$ ) relative to the natural densities in Lake Michigan in this study ( $\sim 10^6 \text{ cells ml}^{-1}$ ), and with different IDM species at a higher temperature, thus making direct comparisons difficult. With respect to experiments on natural bacterioplankton, mixed results have been reported. For river bacterioplankton, short-term mesocosm experiments provided no evidence of a direct feeding effect, while long-term environmental surveys suggested a doubling in bacterioplankton densities (Findlay *et al.*, 1998). In lakes, feeding on bacteria in low nutrient systems was thought to be limited (Cotner *et al.*, 1995), though short-term feeding experiments on natural bacterioplankton from Lake Michigan did detect significant decreases in bacterioplankton densities (Denef *et al.*, 2017).

HNA and LNA populations have been well-characterized in aquatic environments, yet considerable debate remains regarding the characteristics of each population. Initially, it was thought that the HNA population was the active fraction of the bacterial community, whereas the LNA population served as a reservoir of dormant, inactive, dead, dying and damaged cells (Lebaron *et al.*, 2002). Nowadays, the LNA population has been shown to be able to actively grow and to be metabolically active in the environment without adopting HNA properties, such as HNA content and increased cell size (Jochem *et al.*, 2004; Scharek and Latasa, 2007; Wang *et al.*, 2009). Most studies do report an elevated cell-specific activity for the HNA bacteria that can be more than an order of magnitude higher than the activity of the LNA bacteria (Lebaron *et al.*, 2002; Servais *et al.*, 2003). HNA bacteria are also generally larger and exhibit higher growth rates than LNA bacteria (Lebaron *et al.*, 2002; Jochem *et al.*, 2004;

Scharek and Latasa, 2007), and this large, active fraction of the bacterioplankton is preferred by zooplankton grazing (Boenigk *et al.*, 2004; Tadonleke *et al.*, 2005; Garcia-Chaves *et al.*, 2016).

HNA population densities tend to be positively correlated with heterotrophic productivity (Zubkov *et al.*, 2001; Bouvier *et al.*, 2007). Thus, we would expect relatively low HNA densities in Lake Michigan, which has been rendered increasingly oligotrophic (low primary and secondary productivity) since the ingress of IDMs (Evans *et al.*, 2011). HNA population densities in Lake Michigan field samples ( $29.6 \pm 4.2\%$ ,  $n = 30$ ) were even lower than those reported in previous surveys of freshwater lakes ( $40\text{--}42.5\%$ ,  $n = 81$ ) with similar levels of primary productivity ( $2 \pm 1.5$  vs.  $1.5 \pm 1.2 \mu\text{g chlorophyll } a \text{ l}^{-1}$  in our 2015 survey of Lake Michigan) (Bouvier *et al.*, 2007; Shuchman *et al.*, 2013). We observed that IDMs predominantly feed on the HNA population; this may in part explain these lower than expected HNA densities, as Lake Michigan is characterized lake-wide by high densities of IDMs (Nalepa *et al.*, 2010). With the HNA bacteria leveraging as much as 80% of the community's secondary production and mediating up to 70% of the bacterial carbon flux in other aquatic environments, the selective feeding by IDMs may have a significant impact on elemental cycling in lake systems (Zubkov *et al.*, 2001; Scharek and Latasa, 2007).

Community characterization of HNA and LNA populations has shown that there can be significant differences in community composition, with few shared taxa between the populations (Schattenhofer *et al.*, 2011; Vila-Costa *et al.*, 2012). As a consequence, selective removal of a single population (e.g., HNA) will alter the community diversity, which was observed during this filter feeding experiment. Other studies using basic molecular fingerprinting techniques did not observe distinct community structures (Servais *et al.*, 2003; Longnecker *et al.*, 2005). Hence, several scenarios have been developed to explain the dichotomy between HNA and LNA bacterioplankton populations (Bouvier *et al.*, 2007). Our results only allow us to support the scenario in which each population has a separate community structure, since considerable repopulation of the HNA population through growth or potential recruitment from the LNA population can take several days (Gasol *et al.*, 1999; Sintes and del Giorgio, 2014; Baltar *et al.*, 2016).

While few investigations into the impacts of IDM on bacterial community composition have been performed, our observations are congruent with studies that have shown altered composition in the sediment (Frischer *et al.*, 2000; Lohner *et al.*, 2007; Lee *et al.*, 2015) and water column (Denef *et al.*, 2017) following IDM introduction. In these studies, specific taxonomic groups were shown to become relatively enriched within the microbial community. Among others, taxa of the phylum Actinobacteria and the genus



*Polynucleobacter*, which are known to possess LNA-type characteristics such as small cell sizes (Wang *et al.*, 2009), increased in relative abundance during short-term microcosm experiments (Denef *et al.*, 2017). Other taxa (e.g., *Chloroflexi*) became enriched despite their larger cell size, suggesting that multiple phenotypic traits beyond mere cell size determine the feeding success on bacterial taxa. Relative enrichments do need to be interpreted with care as these can provide biased interpretations of the actual taxon abundance dynamics (Nayfach and Pollard, 2016; Props *et al.*, 2016b; Stämmler *et al.*, 2016).

In conclusion, we have shown that advanced data analysis of FCM data can lead to robust predictions of taxonomic diversity within a large dynamic range. We further demonstrated that the diversity of natural bacterioplankton communities can be reliably tracked during sensitive ecological processes in a fast, non-invasive manner. Using this approach we were able to detect subtle shifts in biodiversity emerging within one hour of feeding by IDMs. The selective removal of HNA bacteria was shown to be underlying cause of the loss of biodiversity, suggesting size-selective feeding behaviour in the micrometre range. As a result, IDMs are capable of locally reducing the diversity and productivity of the bacterioplankton community during feeding. The approach presented here can be readily applied to help address a broad range of questions in marine and freshwater systems, for which new analytical and computational tools are needed (Labate *et al.*, 2016). FCM is now also increasingly being developed for other environments such as soils, sediments and sludges, opening new possibilities for these systems as well (Frossard *et al.*, 2016).

## Experimental procedures

### 16S rRNA gene amplicon sequencing analysis

We used a combination of a previously published data set from 2013 and newly generated V4 16S rRNA gene amplicon sequences from 2014 and 2015 lake surveys (see data availability section). V4 amplicon sequencing data from Lake Michigan (2015 survey) and Muskegon Lake (2014, 2015 surveys) were generated exactly as previously described (Schmidt *et al.*, 2016). Samples were taken in September (Fall), April (Spring) and July (Summer). The DNA was extracted according to a previously optimized protocol (McCarthy *et al.*, 2015) and submitted for sequencing of the V4 hypervariable region (515F/806R) by Illumina MiSeq with v2 chemistry (2 × 250 bp). All raw sequencing reads from these surveys were processed together, after which the samples with matching FCM data were extracted.

Contigs were created by merging paired-end reads based on the Phred quality score heuristic (Kozich *et al.*, 2013) in MOTHUR (v.1.38, seed = 777) (Schloss *et al.*, 2009). Contigs were aligned to the Silva database (v123) and filtered from those with (i) ambiguous bases, (ii) more than 8 homopolymers, (iii) a length outside of the 240–275 nt range and (iv)

those not corresponding to the V4 region. The aligned sequences were filtered and dereplicated, and sequencing errors were removed using the *pre.cluster* command. Chimera removal was performed by UCHIME. Sequences were clustered into operational taxonomic units (OTUs) at 97% similarity with the *cluster.split* command (average neighbor algorithm). Sequences were then classified using the TaxAss pipeline (<https://github.com/McMahonLab/TaxAss>), which classifies sequences according to both a small, manually curated freshwater taxonomy database (Newton *et al.*, 2011) and a large, general database (SILVA v123). The complete workflow is available at [https://github.com/rprops/Mothur\\_oligo\\_batch](https://github.com/rprops/Mothur_oligo_batch) and was run in batch mode. For comparison to the FCM data, only the samples comprising the bacterioplankton fraction (0.22–3 µm fraction) were used in further analyses, as this fraction was the most directly comparable to the measurements taken by the flow cytometer.

The cooling water reference data contain publicly available V3–V4 16S rRNA gene amplicon sequences and are available from the NCBI Sequence Read Archive (SRA) under accession number SRP066190. We utilized the OTU-table from a previous publication as basis for the diversity calculations (Props *et al.*, 2016b). This OTU-table was generated according to the same pipeline as described earlier.

### FCM analysis

A total of 1 ml of unfiltered water samples were fixed with 5 µl glutaraldehyde (20%, vol/vol, stock), incubated for 10 min in the dark and flash frozen in liquid nitrogen (storage at –80°C). Prior to FCM analysis, batches of eight samples were sequentially defrosted, acclimated to room temperature, diluted twofold in triplicate and stained with SYBR Green I (10,000× in DMSO; Invitrogen) to a final concentration of 1× SYBR Green I. Samples were then incubated at 37°C for 20 min in the dark and analysed directly on a BD Accuri C6 cytometer (BD Biosciences, Erembodegem, Belgium) in fixed volume mode (50 µl) (Props *et al.*, 2016a). This resulted in a multi-parametric description of each microbial cell by four fluorescence parameters (FL1: 533/30 nm, FL2: 585/40 nm, FL3: > 670 nm long pass, FL4: 675/25 nm) and two scatter parameters (FSC and SSC). Instrument performance was verified daily using eight peak rainbow particles (Spherotech, Lake Forest, IL).

### Phenotypic diversity analysis

The alpha diversities for both the FCM and sequencing data were assessed by the Hill diversity numbers, which incorporate both richness and evenness components (Hill, 1973). We followed the previously published protocol available here: [https://github.com/rprops/Phenoflow\\_package/wiki/1.-Phenotypic-diversity-analysis](https://github.com/rprops/Phenoflow_package/wiki/1.-Phenotypic-diversity-analysis) (Props *et al.*, 2016a). Raw flow cytometer data were exported in FCS format and imported into R (v3.3.0), using functions from the *flowCore* package (v1.38.2). The data were denoised from (in)organic noise based on previous experience with freshwater communities and according to published guidelines for robust denoising (Prest *et al.*, 2013) (as described in Supporting Information Fig. S4). The denoising strategy remained the same for all samples. Samples with less than

10,000 cells were discarded since sample sizes larger than this threshold were required for the robust estimation of the diversity (Supporting Information Fig. S6). All single-cell parameters were normalized based on the maximum signal height (-H) of the FL1 parameter. The *Diversity* function from the *Phenoflow* package (v1.0, [https://github.com/rprops/Phenoflow\\_package](https://github.com/rprops/Phenoflow_package)) was then used to calculate the phenotypic alpha diversities of the four primary parameters (FL1-H, FL3-H, FSC-H and SSC-H). Errors on the diversities were generated after 100 bootstraps and propagated to the mean diversity over the three technical replicates. The kernel density estimations were performed with a bandwidth of 0.01, a grid size of  $128 \times 128$  and a rounding factor of 3. The alpha diversity was evaluated through the first three Hill numbers:  $D_0$ ,  $D_1$  and  $D_2$ , which correspond to the observed richness, the exponential of Shannon entropy and the inverse Simpson index respectively. Beta diversity analyses were performed by principal coordinate analysis (PCoA) of the phenotypic fingerprints (*flow-Basis* function,  $d = 3$ ,  $bw = 0.01$ ) using the Bray–Curtis dissimilarity metric (*beta\_div\_fcm* function, default settings). Contrasts between the phenotypic fingerprints of the control and treatment groups were made by the *fp\_contrasts* function (see tutorial here: [https://github.com/rprops/Phenoflow\\_package/wiki/3.-Making-contrasts](https://github.com/rprops/Phenoflow_package/wiki/3.-Making-contrasts)).

#### Taxonomic diversity analysis

For calculating the taxonomic alpha diversity, we used the *Diversity\_16S* function from the *Phenoflow* package because this allowed a direct comparison between the taxonomic and phenotypic diversity metrics. The community data were not rarefied because our hypothesis was that the taxonomic diversity was correlated with an unrelated variable, the phenotypic diversity. Subsampling to the lowest sample size would result in the poorest estimate of the taxonomic diversity for all samples, thereby potentially obscuring the true relationship between these variables. Instead, we selected only samples which had a sample size larger than 10,000 reads ( $n = 138$ ), generated 100 bootstrap samples for each sample and took the average diversity as the sample representative diversity. Parallel results of our analyses for the rescaled data to 10,000 reads are available in Supporting Information Fig. S5. For beta diversity analysis, the OTU abundances were rescaled by calculating their proportions and multiplying them by the minimum sample size present in the data set (McMurdie and Holmes, 2014). The beta diversity was then assessed by PCoA of the Bray–Curtis dissimilarity matrix, which was calculated based on the taxon proportions instead of the read counts in order to be directly comparable to the phenotypic beta diversity approach.

#### Feeding experiment

Quagga mussels and lake water (5 m above lake floor) were collected at 45 m deep from Lake Michigan ( $43^{\circ}12'N$ ,  $86^{\circ}27'W$ ). Mussels were rinsed of adhering sediment and were transported (< 8 h) submerged in lake water at  $5-7^{\circ}C$ . The standard handling and experimental design of (Vanderploeg *et al.*, 2010) was followed and is briefly outlined as follows. In the lab, the mussels were cleaned of debris and placed in a tank filled with 90 l of 153  $\mu m$ -screened Lake

Michigan water in order to remove grazing mesozooplankton in an environmental room set to ambient temperature of the lake water ( $9.7^{\circ}C$ ). The next morning the mussels were transferred to a 40 l aquarium with 153  $\mu m$  screened Lake Michigan water for 2 h. The mussel cleaning and  $\sim 14$  h re-acclimation period allowed the removal of external periphyton and debris, cleared the mussel guts of sediment ingested during capture and gave mussels time to reach digestive equilibrium with their natural food source. All materials were washed with bleach and rinsed with deionized water to minimize bacterial contamination. Seven 19 l HDPE cylindrical containers were filled with 12 l of 153  $\mu m$ -screened lake water each. Forty-five adult mussels were spread evenly across three containers (15 mussels per container) and two containers remained mussel-free. Gentle mixing was provided by bubbling air through a pipette, and all experiments were carried out under dim light ( $\sim 8 \mu mol \text{ quanta } m^{-2} s^{-1}$ ). Water samples were taken before the addition of the mussels and every 0.5 h after the mussels showed signs of active feeding (after approx. 15 min). The number of mussels added and the experiment duration were chosen to allow healthy mussels to clear 30–60% of preferred seston. As shown in a previous study (Deneff *et al.*, 2017), our procedure ensures that mussel-associated bacteria do not significantly impact observed shifts in bacterial community composition over the duration of the experiment. One millilitre water samples from the top water layer were taken every 30 min throughout and at the end of the 3 h experiment. The samples were fixed with 5  $\mu l$  glutaraldehyde (20%, vol/vol), incubated for 10 min in the dark and flash frozen in liquid nitrogen (storage at  $-80^{\circ}C$ ).

#### Statistical analysis

All statistical analyses were performed in the R statistical environment (v3.3.0) (R Core Team, 2015), using functions from the *vegan* (v2.4-1), *sandwich* (v2.3-4), *MASS* (v7.3-45), *car* (v2.1-3), *phyloseq* (v1.16.2), *lme4* (v0.9-34) and *caret* (v6.0-73) packages. Errors on all summary statistics represent standard deviations on the mean and were calculated by propagating individual standard deviations as randomly distributed, independent errors. Ordinary least squares regression was used to relate the phenotypic diversity to the taxonomic diversity (both  $\log_2$  transformed). Model assumptions (i.e., normality and homoscedasticity) were evaluated through analysis of the residuals (Supporting Information Fig. S2). Goodness-of-fit statistics were calculated through tenfold cross validation with 100 repeats. Inference on the temporal treatment effect on the phenotypic diversity ( $D_2$ ) was performed by spline regression. We opted for natural splines because these provide more stable estimates at the boundaries (James *et al.*, 2014). Splines were given 3 degrees of freedom, allowing two knots to occur at the 33.3% and 66.6% quantiles (i.e., at time points 1 and 2 h). Parameter estimation was performed by the robust ordinary least squares method. Robust linear regression is a variation on the traditional ordinary least squares (ols) regression that provides more correct inference when assumptions for ols regression are invalid (i.e., less sensitive to outliers). Due to the presence of moderate temporal autocorrelation in the model residuals (Supporting Information Fig. S8), robust parameter errors, calculated from the autocorrelation adjusted covariance matrix (*vcovHAC*

function), were used in the statistical inference (Wald test). Differences between groups in the beta diversity analysis were evaluated by means of permutational multivariate ANOVA (PERMANOVA, *adonis* function, 999 permutations) of the Bray–Curtis dissimilarity matrix, after confirmation of the homogeneity of the variance in the groups (*betadisper* function). Similarity between beta diversity analyses was evaluated through Procrustes analysis (*protest* function, 999 permutations). Temporal trends in the HNA cell density, as well as the feeding rate, were determined through robust ordinary least squares linear regression. Statistical inference on the model parameters was performed with the Wald test. The clearance rate (CR) was determined based on the robust linear regression of the HNA cell dynamics:

$$CR = \frac{V}{n} \times \frac{a}{b}$$

$V$  is the water volume of the container (ml),  $n$  is the average dry weight of the mussels (mg),  $a$  is the slope of the regression (cells ml<sup>-1</sup> h<sup>-1</sup>) and  $b$  is the intercept of the regression (cells ml<sup>-1</sup>).

#### Data availability

The entire data-analysis pipeline is available as an R Markdown document at [https://github.com/DenefLab/EnvMicro\\_Props2017](https://github.com/DenefLab/EnvMicro_Props2017). Raw FCM data are available on FlowRepository under accession IDs FR-FCM-ZZNA (cooling water), FR-FCM-ZYZA (mussel feeding experiment) and FR-FCM-ZYZN (Lake Michigan and Muskegon Lake survey). Newly generated V4 16S rRNA sequences from lake Michigan and Muskegon Lake were deposited on the NCBI SRA under BioProject ID PRJNA412984 and PRJNA412983, respectively. For Lake Michigan, the 2013 data set is publicly available on the Joint Genome Institute's genome data portal (<http://genome.jgi.doe.gov/>; project IDs 1041195 and 1041198).

#### Acknowledgements

We thank Paul Glyshaw for assistance in the sampling and transport of the mussels and Thomas Johengen for help in setting up the experiments. We thank the crew of the R/V W.G. Jackson and Dr. Bopaiah Biddanda for enabling sampling on Muskegon Lake, the crew on the R/V Laurentian for sampling on Lake Michigan and Edna Chiang, Kyle Buffin and Amadeus Twu for assistance with DNA extractions of the Muskegon Lake samples. MLS is supported by the National Science Foundation Graduate Research Fellowship Program. Part of this work was supported by the Community Sequencing Program (U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, supported under Contract No. DE-AC02-05CH11231). This is GLERL Contribution No. 1869. This work was supported through the Inter-University Attraction Pole (IUAP) 'I-manager' funded by the Belgian Science Policy (BELSPO, P7/25) and Geconcentreerde Onderzoeksactie (GOA) from Ghent University (BOF15/GOA/006). RP was supported by Ghent University (BOFDOC2015000601) and a Sofina Gustave-Boël grant from the Belgian American Educational Foundation. The authors declare that there exist no conflicts of interest.

#### Author contributions

RP and VJD wrote the paper. RP performed laboratory work and all statistical analyses. HAV contributed in the laboratory work. JH performed the FCM data acquisition. MLS generated sequencing data. VJD, HAV, NB and RP designed the study. All authors reviewed and approved the manuscript.

#### References

- Baltar, F., Palovaara, J., Unrein, F., Catala, P., Hornak, K., Simek, K., *et al.* (2016) Marine bacterial community structure resilience to changes in protist predation under phytoplankton bloom conditions. *ISME J* **10**: 568–581.
- Boenigk, J., Stadler, P., Wiedroither, A., and Hahn, M.W. (2004) Strain-specific differences in the grazing sensitivities of closely related ultramicrobacteria affiliated with the *Polynucleobacter* cluster. *Appl Environ Microbiol* **70**: 5787–5793.
- Bouvier, T., del Giorgio, P.A., and Gasol, J.M. (2007) A comparative study of the cytometric characteristics of high and low nucleic-acid bacterioplankton cells from different aquatic ecosystems. *Environ Microbiol* **9**: 2050–2066.
- Chapin, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., *et al.* (2000) Consequences of changing biodiversity. *Nature* **405**: 234–242.
- Cotner, J.B., Gardner, W.S., Johnson, J.R., Sada, R.H., Cavaletto, J.F., and Heath, R.T. (1995) Effects of zebra mussels (*Dreissena polymorpha*) on bacterioplankton: evidence for both size-selective consumption and growth stimulation. *J Great Lakes Res* **21**: 517–528.
- Datta, M.S., Sliwerska, E., Gore, J., Polz, M.F., and Cordero, O.X. (2016) Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nat Commun* **7**.
- Denef, V.J., Carrick, H.J., Cavaletto, J., Chiang, E., Johengen, T.H., and Vanderploeg, H.A. (2017) Lake bacterial assemblage composition is sensitive to biological disturbance caused by an invasive filter feeder. *mSphere* **2**.
- Evans, M.A., Fahnenstiel, G., and Scavia, D. (2011) Incidental oligotrophication of North American Great Lakes. *Environ Sci Technol* **45**: 3297–3303.
- Faust, K., Lahti, L., Gonze, D., de Vos, W.M., and Raes, J. (2015) Metagenomics meets time series analysis: unravelling microbial community dynamics. *Curr Opin Microbiol* **25**: 56–66.
- Findlay, S., Pace, M.L., and Fischer, D.T. (1998) Response of heterotrophic planktonic bacteria to the zebra mussel invasion of the tidal freshwater Hudson River. *Microbial Ecol* **36**: 131–140.
- Finlay, J.C., Sterner, R.W., and Kumar, S. (2007) Isotopic evidence for in-lake production of accumulating nitrate in lake superior. *Ecol Appl* **17**: 2323–2332.
- Frischer, M.E., Nierzwicki-Bauer, S.A., Parsons, R.H., Vathanadorn, K., and Waitkus, K.R. (2000) Interactions between zebra mussels (*Dreissena polymorpha*) and microbial communities. *Can J Fish Aquat Sci* **57**: 591–599.
- Frossard, A., Hammes, F., and Gessner, M.O. (2016) Flow cytometric assessment of bacterial abundance in soils, sediments and sludge. *Front Microbiol* **7**.



- Garcia-Chaves, M.C., Cottrell, M.T., Kirchman, D.L., Ruiz-Gonzalez, C., and del Giorgio, P.A. (2016) Single-cell activity of freshwater aerobic anoxygenic phototrophic bacteria and their contribution to biomass production. *ISME J* **10**: 1579–1588.
- Gasol, J.M., Zweifel, U.L., Peters, F., Fuhrman, J.A., and Hagstrom, A. (1999) Significance of size and nucleic acid content heterogeneity as measured by flow cytometry in natural planktonic bacteria. *Appl Environ Microbiol* **65**: 4475–4483.
- Gunther, S., Hubschmarm, T., Rudolf, M., Eschenhagen, M., Roske, I., Harms, H., and Muller, S. (2008) Fixation procedures for flow cytometric analysis of environmental bacteria. *J Microbiol Methods* **75**: 127–134.
- Hammes, F., and Egli, T. (2010) Cytometric methods for measuring bacteria in water: advantages, pitfalls and applications. *Analyt Bioanal Chem* **397**: 1083–1095.
- Hammes, F., Berney, M., Wang, Y., Vital, M., Koester, O., and Egli, T. (2008) Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. *Water Res* **42**: 269–277.
- Higgins, S.N., and Vander Zanden, M.J. (2010) What a difference a species makes: a meta-analysis of dreissenid mussel impacts on freshwater ecosystems. *Ecol Monogr* **80**: 179–196.
- Hill, M.O. (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**: 427–432.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014) *An Introduction to Statistical Learning: With Applications in R*. Berlin, Germany: Springer Publishing Company, Incorporated.
- Jochem, F.J., Lavrentyev, P.J., and First, M.R. (2004) Growth and grazing rates of bacteria groups with different apparent DNA content in the Gulf of Mexico. *Mar Biol* **145**: 1213–1225.
- Koch, C., Harnisch, F., Schroeder, U., and Mueller, S. (2014) Cytometric fingerprints: evaluation of new tools for analyzing microbial community dynamics. *Front Microbiol* **5**: 1–10.
- Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina Sequencing Platform. *Appl Environ Microbiol* **79**: 5112–5120.
- Labbate, M., Seymour, J.R., Lauro, F., and Brown, M.V. (2016) Anthropogenic impacts on the microbial ecology and function of aquatic environments. *Front Microbiol* **7**: 1044.
- Lebaron, P., Servais, P., Baudoux, A.C., Bourrain, M., Courties, C., and Parthuisot, N. (2002) Variations of bacterial-specific activity with cell size and nucleic acid content assessed by flow cytometry. *Aquat Microb Ecol* **28**: 131–140.
- Lee, P.O., McLellan, S.L., Graham, L.E., and Young, E.B. (2015) Invasive dreissenid mussels and benthic algae in Lake Michigan: characterizing effects on sediment bacterial communities. *FEMS Microbiol Ecol* **91**: 1–10.
- Levine, U.Y., Teal, T.K., Robertson, G.P., and Schmidt, T.M. (2011) Agriculture's impact on microbial diversity and associated fluxes of carbon dioxide and methane. *ISME J* **5**: 1683–1691.
- Lohner, R.N., Sigler, V., Mayer, C.M., and Balogh, C. (2007) A comparison of the benthic bacterial communities within and surrounding *Dreissena* clusters in lakes. *Microb Ecol* **54**: 469–477.
- Longnecker, K., Sherr, B.F., and Sherr, E.B. (2005) Activity and phylogenetic diversity of bacterial cells with high and low nucleic acid content and electron transport system activity in an upwelling ecosystem. *Appl Environ Microbiol* **71**: 7737–7749.
- McCarthy, A., Chiang, E., Schmidt, M.L., and Deneff, V.J. (2015) RNA preservation agents and nucleic acid extraction method bias perceived bacterial community composition. *PLoS One* **10**: e0121659.
- McMurdie, P.J., and Holmes, S. (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol* **10**: e1003531.
- Nalepa, T.F., Fanslow, D.L., and Pothoven, S.A. (2010) Recent changes in density, biomass, recruitment, size structure, and nutritional state of *Dreissena* populations in southern Lake Michigan. *J Great Lakes Res* **36**: 5–19.
- Nayfach, S., and Pollard, K.S. (2016) Toward Accurate and Quantitative Comparative Metagenomics. *Cell* **166**: 1103–1116.
- Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., and Bertilsson, S. (2011) A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* **75**: 14–49.
- Padilla, C.C., Ganesh, S., Gantt, S., Huhman, A., Parris, D.J., Sarode, N., and Stewart, F.J. (2015) Standard filtration practices may significantly distort planktonic microbial diversity estimates. *Front Microbiol* **6**: 547.
- Pires, L.M.D., Jonker, R.R., Van Donk, E., and Laanbroek, H.J. (2004) Selective grazing by adults and larvae of the zebra mussel (*Dreissena polymorpha*): application of flow cytometry to natural seston. *Freshwater Biol* **49**: 116–126.
- Prest, E.I., Hammes, F., Kotzsch, S., van Loosdrecht, M.C.M., and Vrouwenvelder, J.S. (2013) Monitoring microbiological changes in drinking water systems using a fast and reproducible flow cytometric method. *Water Res* **47**: 7131–7142.
- Props, R., Monsieurs, P., Mysara, M., Clement, L., and Boon, N. (2016a) Measuring the biodiversity of microbial communities by flow cytometry. *Methods Ecol Evol* **7**: 1376–1385.
- Props, R., Kerckhof, F.-M., Rubbens, P., De Vrieze, J., Hernandez Sanabria, E., Waegeman, W., et al. (2016b) Absolute quantification of microbial taxon abundances. *ISME J* **11**: 584–587.
- R Core Team (2015) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Scharek, R., and Latasa, M. (2007) Growth, grazing and carbon flux of high and low nucleic acid bacteria differ in surface and deep chlorophyll maximum layers in the NW Mediterranean Sea. *Aquat Microb Ecol* **46**: 153–161.
- Schattenhofer, M., Wulf, J., Kostadinov, I., Glockner, F.O., Zubkov, M.V., and Fuchs, B.M. (2011) Phylogenetic characterisation of picoplanktonic populations with high and low nucleic acid content in the North Atlantic Ocean. *Syst Appl Microbiol* **34**: 470–475.
- Schimel, J.P., and Gullede, J. (1998) Microbial community structure and global trace gases. *Glob Change Biol* **4**: 745–758.



- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Schmalenberger, A., Schwieger, F., and Tebbe, C.C. (2001) Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in PCR-based microbial community analyses and genetic profiling. *Appl Environ Microbiol* **67**: 3557–3563.
- Schmidt, M.L., White, J.D., and Deneff, V.J. (2016) Phylogenetic conservation of freshwater lake habitat preference varies between abundant bacterioplankton phyla. *Environ Microbiol* **18**: 1212–1226.
- Servais, P., Casamayor, E.O., Courties, C., Catala, P., Parthuisot, N., and Lebaron, P. (2003) Activity and diversity of bacterial cells with high and low nucleic acid content. *Aquat Microb Ecol* **33**: 41–51.
- Shade, A., Peter, H., Allison, S.D., Baho, D.L., Berga, M., Burgmann, H., *et al.* (2012) Fundamentals of microbial community resistance and resilience. *Front Microbiol* **3**.
- Shuchman, R.A., Sayers, M., Fahnenstiel, G.L., and Leshkevich, G. (2013) A model for determining satellite-derived primary productivity estimates for Lake Michigan. *J Great Lakes Res* **39**: 46–54.
- Silverman, H., Achberger, E.C., Lynn, J.W., and Dietz, T.H. (1995) Filtration and utilization of laboratory-cultured bacteria by *Dreissena polymorpha*, *Corbicula fluminea*, and *Carunculina texensis*. *Biol Bull* **189**: 308–319.
- Singh, B.K., Bardgett, R.D., Smith, P., and Reay, D.S. (2010) Microorganisms and climate change: terrestrial feedbacks and mitigation options. *Nat Rev Microbiol* **8**: 779–790.
- Sintes, E., and del Giorgio, P.A. (2014) Feedbacks between protistan single-cell activity and bacterial physiological structure reinforce the predator/prey link in microbial food-webs. *Front Microbiol* **5**: 453.
- Stammler, F., Glasner, J., Hiergeist, A., Holler, E., Weber, D., Oefner, P.J., *et al.* (2016) Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. *Microbiome* **4**: 28.
- Tadonleke, R.D., Planas, D., and Lucotte, A. (2005) Microbial food webs in boreal humic lakes and reservoirs: ciliates as a major factor related to the dynamics of the most active bacteria. *Microb Ecol* **49**: 325–341.
- Tang, H.J., Vanderploeg, H.A., Johengen, T.H., and Liebig, J.R. (2014) Quagga mussel (*Dreissena rostriformis bugensis*) selective feeding of phytoplankton in Saginaw Bay. *J Great Lakes Res* **40**: 83–94.
- Vanderploeg, H.A., Liebig, J.R., Nalepa, T.F., Fahnenstiel, G.L., and Pothoven, S.A. (2010) *Dreissena* and the disappearance of the spring phytoplankton bloom in Lake Michigan. *J Great Lakes Res* **36**: 50–59.
- Vanderploeg, H.A., Nalepa, T.F., Jude, D.J., Mills, E.L., Holeck, K.T., Liebig, J.R., *et al.* (2002) Dispersal and emerging ecological impacts of Ponto-Caspian species in the Laurentian Great Lakes. *Can J Fish Aquat Sci* **59**: 1209–1228.
- Vila-Costa, M., Gasol, J.M., Sharma, S., and Moran, M.A. (2012) Community analysis of high- and low-nucleic acid-containing bacteria in NW Mediterranean coastal waters using 16S rDNA pyrosequencing. *Environ Microbiol* **14**: 1390–1402.
- Wang, Y., Hammes, F., Boon, N., Chami, M., and Egli, T. (2009) Isolation and characterization of low nucleic acid (LNA)-content bacteria. *ISME J* **3**: 889–902.
- Willis, A., and Bunge, J. (2015) Estimating diversity via frequency ratios. *Biometrics* **71**: 1042–1049.
- Yu, Z.T., and Morrison, M. (2004) Comparisons of different hypervariable regions of rrs genes for use in fingerprinting of microbial communities by PCR-denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **70**: 4800–4806.
- Zhou, J.Z., Xue, K., Xie, J.P., Deng, Y., Wu, L.Y., Cheng, X.H., *et al.* (2012) Microbial mediation of carbon-cycle feedbacks to climate warming. *Nat Clim Change* **2**: 106–110.
- Zubkov, M.V., Fuchs, B.M., Burkill, P.H., and Amann, R. (2001) Comparison of cellular and biomass specific activities of dominant bacterioplankton groups in stratified waters of the Celtic Sea. *Appl Environ Microbiol* **67**: 5210–5218.

### Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Fig. S1.** Validation for the use of the phenotypic diversity (derived from FCM) across environments with varying degrees of taxonomic diversity (derived from 16S rRNA gene amplicon sequencing,  $n = 138$ ). The cooling water samples represent bacterioplankton communities sampled throughout two 40-day temporal surveys of a cooling water system of a nuclear test reactor (Props *et al.*, 2016a). Lake Michigan and Muskegon lakes samples represent bacterioplankton communities sampled over a productivity gradient, at various depths (110, 45 and 15 m) and throughout three seasons (Fall, Spring and Summer). Fall, Spring and Summer denote samples taken in September, April and July respectively. The shaded area represents the 95% confidence interval around the ordinary least squares regression model. Both diversities are depicted on a log2 scale. Bootstrap error intervals fell within the label size and were not displayed.

**Fig. S2.** Residual analysis for the ordinary least squares regression between the taxonomic and phenotypic diversities of order 2 ( $D_2$ ). Left panel: quantile–quantile plot indicating approximate normality of the residuals. 95% confidence intervals were calculated on 10,000 bootstraps. Middle panel: Pearson residuals plotted against the predicted values indicate homoscedastic errors. Right panel: observed values plotted against predicted values.

**Fig. S3.** Ordinary least squares regression between the phenotypic diversity and taxonomic diversity of orders 0 (left) and 1 (right) ( $D_0$ ,  $D_1$ ). In addition to the variance explained by the model (adj.  $r^2$ ), the Pearson's correlation coefficient has also been provided ( $r_p$ ).

**Fig. S4.** Gating strategy for denoising the raw flow cytometry data visualized for the control and treatment samples at three time points. Fluorescence values were arcsinh-transformed and were normalized by the maximum fluorescence intensity. The outer boundary encompasses the total microbial community, while the inner FL1-H threshold (FL1-H = 0.67) separates the high nucleic acid (HNA) populations from the low nucleic acid (LNA) population. Color intensity is proportional to the log-scaled density value.

**Fig. S5.** Regression analysis between the taxonomic and phenotypic diversity of orders 0 ( $D_0$ ), 1 ( $D_1$ ) and 2 ( $D_2$ ) after rescaling sample sizes to 10,000 reads.

**Fig. S6.** Evaluation of required sample size for robust calculation of the phenotypic diversity illustrated for one randomly chosen sample (control microcosm at time point 1 h). The phenotypic diversity was calculated for 100 bootstrap samples at each specified sample size (number of cells). Parameter settings were kept identical to those used in the other analyses.

**Fig. S7.** Autocorrelation plots for the smoothing spline regressions of the phenotypic diversity of order 2 ( $D_2$ ). Blue dashed lines indicate thresholds for significant autocorrelation between the temporal profiles of the Pearson residuals, separated by the specified lag interval.

**Table S1.** Multiple comparisons between slopes of linear regression models between taxonomic (16S) and phenotypic diversity (FCM) for each environment (see Supporting Information Fig. S1 for visualization of each regression).